

ElasticBLAST

Christiam Camacho, Greg Boratyn, Anatoliy Kuznetsov, Victor Joukov, Yuri Merezhuk, Yan Raytselis, Tom Madden, Vadim Zalunin



U.S. National Library of Medicine
National Center for Biotechnology Information



What is ElasticBLAST?

- It is a command line tool to
- facilitate aligning large volumes of nucleotide and protein sequences
- against popular BLAST databases.
 - These cannot be done in WebBLAST or easily with BLAST+

Currently under active development

ElasticBLAST application

- Annotation of mRNA transcripts with functional and biological processes
- BLAST is used to align an assembled transcriptome against annotated databases of proteins or nucleotides.
 - blastp open reading frames against non-redundant database (nr)
 - rpsblast open reading frames against Conserved Domain Database (cdd)
 - blastn transcripts against nucleotide database (nt)
 - rpstblastn transcripts against CDD database
- Example annotation project:
 - *Opuntia streptacantha* (nopal) transcriptome ([PRJNA320545](#)): 474,563 transcripts (217 million base pairs)



Components and key features

- Docker BLAST+

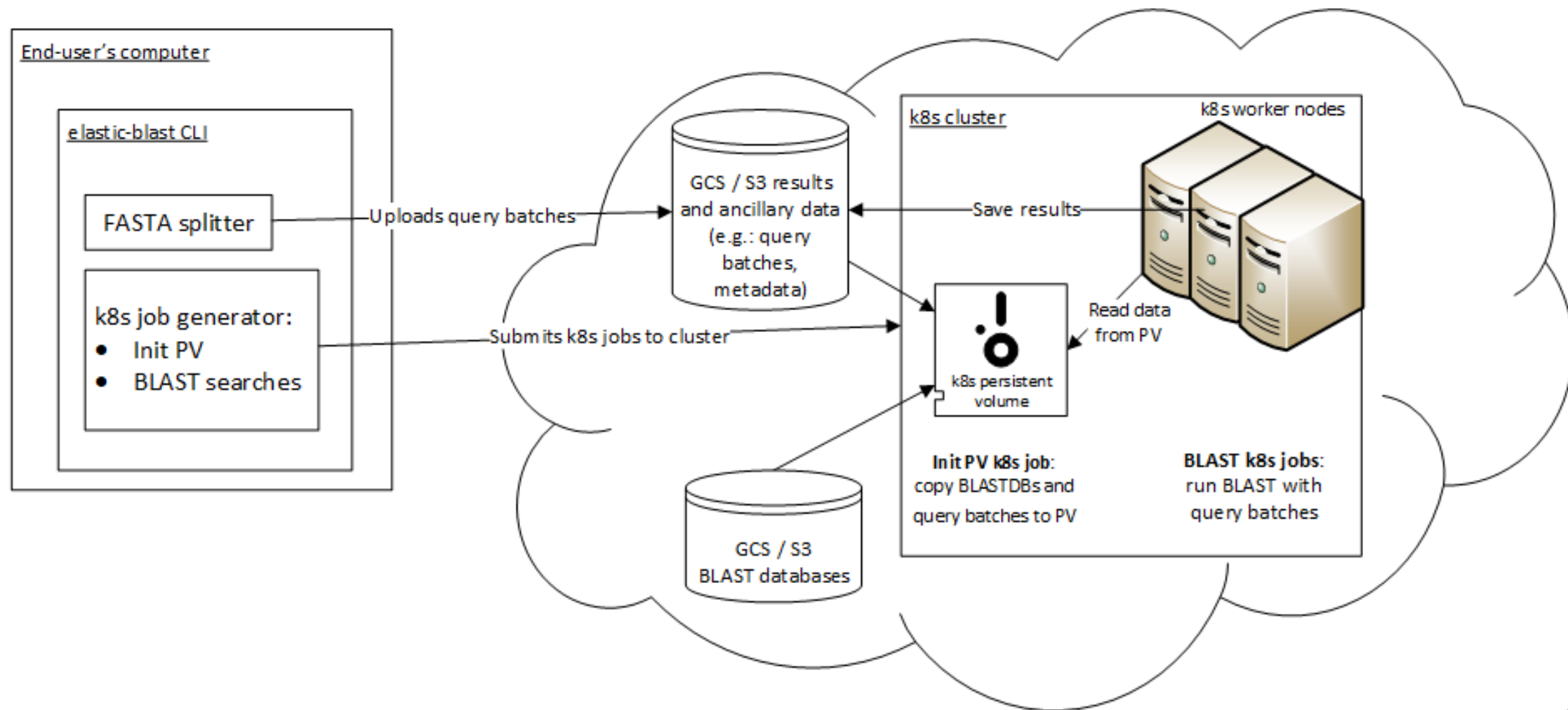


- BLAST databases in the cloud

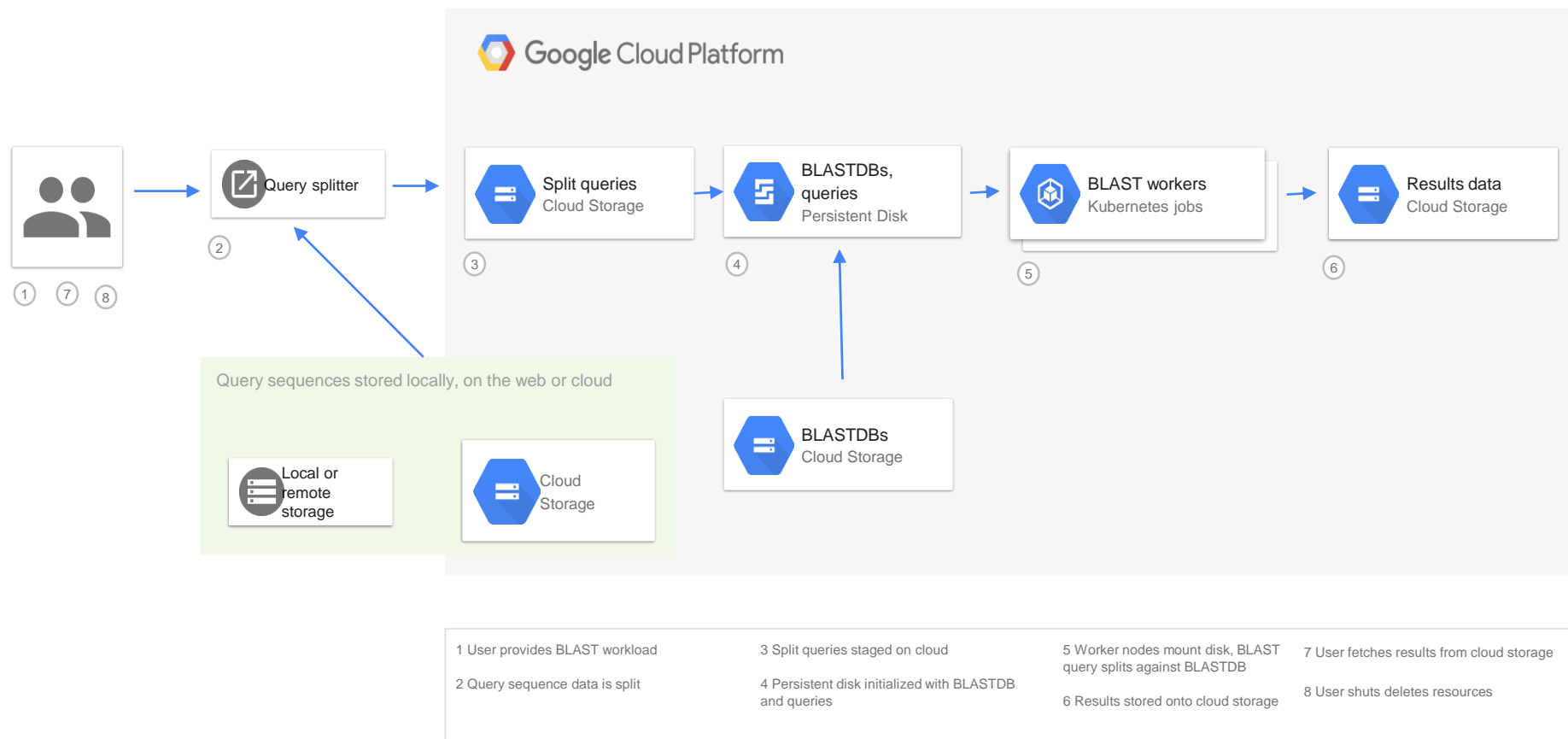


- Leverage the ability to easily procure computational power and storage in the cloud
- Pay-as-you-go model: you only pay for what you use

ElasticBLAST architecture



ElasticBLAST Work Flow



ElasticBLAST Performance example

- Test case:
 - Megablast, default parameters
 - All hepatitis B virus sequences (87k seqs, 80 million bases)
 - nt database (56 million seqs, 256 billion bases)
- On a single 32-vCPU GCP machine
 - Cost: ~\$7
 - Runtime: ~4 hrs
- On 5-machine k8s cluster (160 vCPUs, 600GB RAM)
 - Cost: ~\$7
 - Runtime: ~55 minutes

Want to try ElasticBLAST?

- NCBI is looking for alpha testers
- Interested? Please Contact Dave Arndt at arndtdj@nih.gov

Thank you.

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

Greg Boratyn

Victor Joukov

Yuri Merezhuik

Yan Raytselis

Anatoliy Kuznetsov

Tom Madden

Vadim Zalunin

Roberto Vera Alvarez

Dave Arndt

Scott McGinnis

Peter Cooper

Wayne Matten

Tao Tao

Valerie Schneider

Rodney Brister

David Landsman

Eugene Yaschenko

Kim Pruitt

Steve Sherry